

Project 1

1. Data: [3, 7, 8, 3, 7, 3, 3, 6, 2, 3, 3, 2, 3, 4, 3, 8, 2, 3, 8, 4, 2, 10, 6, 3, 4, 9, 3, 6, 4, 2, 4, 2, 6, 4, 3, 8, 5, 2, 1, 8, 1, 2, 6, 5, 4, 3, 3, 7, 2, 7, 3, 4, 2, 11, 2, 6, 5, 4, 8, 2, 3, 7, 2, 4, 6, 4, 3, 5, 6, 2, 3, 5, 5, 5, 5, 6, 6, 4, 8, 9, 8, 2, 3, 8, 7, 2, 3, 7, 3, 6, 2, 3, 9, 3, 6, 4, 4, 3, 8, 3, 5, 2, 9, 3, 8, 8, 2, 6, 5, 3, 4, 5, 2, 3, 3, 4, 2, 8, 5, 7, 8, 4, 3, 7, 6, 6, 5, 2, 3, 7, 12, 6, 2, 5, 6, 5, 6, 2, 5, 2, 3, 1, 7, 7, 3, 5, 4, 4, 1, 6, 4]

2. Summary Stats: [4.6158940397351, 4.0, 3, 2.281108500000038, 5.203455988772423, 11, 1, 12, 697, 151]

3. The two Adams summaries differ quite a bit from each other. If both of the Adams letters were anonymous, I think it would be very difficult to identify them solely based on the summary stats. Overall, the word length analysis does not do enough to resolve disputed documents.

4. I think it would be useful to know each of these authors specific writing styles and see if there is a way to quantify it. Like if one writer used commas more often than others. A more defining characteristic than word lengths, in my opinion.